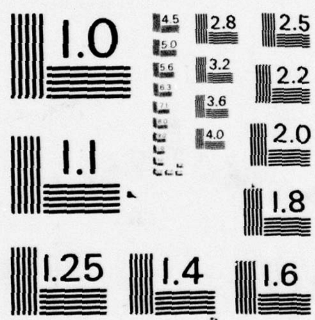
AD-A033 316 RICE UNIV HOUSTON TEX DEPT OF ELECTRICAL ENGINEERING F/G 12/1
SOME DISTANCE MEASURES AND THEIR USE IN FEATURE SELECTION.(U)
1976 P PAPANTONI-KAZAKOS AF-AFOSR-3156-77
UNCLASSIFIED AFOSR-TR-76-1231 NL

1 OF 1
AD
A033316

END
DATE
FILMED
2-77



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(18) AFOSR - TR-76-[redacted] 1231 (19)

(5)

(6)

SOME DISTANCE MEASURES AND THEIR USE IN FEATURE SELECTION.

(9) Interim rept.,

(10) by P. Papantoni-Kazakos
Electrical Engineering
Rice University

(11) 1976

(12) 37p.

Abstract

The Bhattacharyya, I-divergence, Vasershtein, variational and Lévy distances are evaluated, compared and used for the reduction of n data to one feature. This reduction is obtained through a restricted linear transformation and the original data are assumed to be originating from two different jointly Gaussian classes.

It is found that the Bhattacharyya, I-divergence and Vasershtein distances give the same "optimal" linear transformation that applied on the original n data result in one feature with maximum possible distance between classes.

The distortion measures considered in the Vasershtein distance are $|x-y|$ and $(x-y)^2$. For the same distance measures and classes with equal covariances the Lévy distance results in the same "optimal" linear transformation.

(15) AF-AFOSR-3156-77

(16) 2304

(17) A5

This work was supported by the Air Force Grant AFOSR 77-3156

1473
403244

I. Introduction

Distances between probability measures have been used as valuable evaluation and optimization criteria in signal selection [8], pattern recognition [22], universal coding [16, 18], and statistical robustness [21].

The choice of the proper distance must be the result of a carefully weighted decision, where both the calculability and the power of the distance are considered.

An additional factor that is important for the choice of the proper distance in statistically ill-defined environments is the sensitivity of the distance to the changes of the underlined measures.

In section 3 of this paper some comparative evaluation between Bhattacharyya, I-divergence, Vasershtein, variational and Lévy distances is presented.

In section 4, the above distances are used for the linear reduction of n data coming from two Gaussian populations, to one feature.

2. Preliminaries

Let $(\Omega, \mathcal{G}, \mu)$ be a separable, complete measure space, where $\mathcal{G}$ is a σ-algebra of sets in Ω and μ is the measure. Let us now define two different measures μ_1, μ_2 in $(\Omega, \mathcal{G})$ such that $\mu_1(\Omega) = \mu_2(\Omega)$.

If A,B are set-members of the σ-algebra $\mathcal{G}$, and if ρ (A,B) is a penalty function, or distortion measure defined on $\mathcal{G}$, then the Prokhorov and Vasershtein distances between the two measures μ_1, μ_2 are defined as follows:

Prokhorov distance:

$$d_{P,\rho}(\mu_1,\mu_2) \quad \inf\{\epsilon : \mu_1(A) \le \mu_2(\cup B; B: B\in\mathcal{G},\ \rho(A,B) \le \epsilon) + \epsilon,$$

$$\mu_2(A) \le \mu_1(\cup B: B\in\mathcal{G}, \rho(A,B) \le \epsilon) + \epsilon; \forall A\in\mathcal{G}\} \qquad (1)$$

Vasershtein distance:

$$d_{V,\rho}(\mu_1,\mu_2) \quad \inf_{\substack{\text{all joint} \\ \text{measures } r(A,B):\mu_1(A),\mu_2(B) \text{ marginals}}} E_{r(.,.)}\{\rho(A,B)\} \tag{2}$$

The distortion measure considered is incorporated in the definition of the Prokhorov and Vasershtein distances. Also, the Prokhorov distance is very sensitive to the changes of the μ_1, μ_2 measures, while for some choices of distortion measure $\rho(.,.)$, the Vasershtein distance is only a function of second order statistical characteristics (as shown in [16] and discussed in section 3). Finally, while the Prokhorov distance is bounded from above by $\mu_1(\Omega)\ \mu_2(\Omega)$, the Vasershtein distance is bounded only if the distortion measure $\rho(A,B)$ is, for all $A,B \in \mathcal{G}$.

If the distortion measure $\rho(A,B)$ is a metric and is such that $\rho(A,A)=0$, then it was shown by Dobrushin ([13] pp 496) that $d_{V,\rho}(\mu_1,\mu_2)$ is also a metric and $d_{V,\rho}(\mu_1,\mu_2)=0$ if and only if $\mu_1 \equiv \mu_2$. The Prokhorov distance has the same properties for any $\rho(.,.)$ choice. In Dobrushin the relationship:

$$d_{V,\rho}(\mu_1,\mu_2) \geq \left[d_{P,\rho}(\mu_1,\mu_2)\right]^2 \tag{3}$$

is also stated. A formal proof of this relationship is given in appendix A.

If the space Ω is a Euclidean space, and f_1, f_2 are cumulative distributions, the Prokhorov distance reduces to the Lévy distance. For Ω the x-real line, and for μ_1,μ_2 being the one-dimensional cumulative distributions $F_1(x), F_2(x)$ correspondingly, the Lévy distance is given by the following expression:

$$d_{L,\rho}(F_1,F_2) \quad \inf\left\{\varepsilon : F_2(\min(y:\rho(x,y)\leq\varepsilon)) - \varepsilon \leq F_1(x) \leq F_2(\max(y:\rho(x,y))\leq\varepsilon) + \varepsilon ; \forall x\right\} \tag{4}$$

The Lévy distance has the same properties with the Prokhorov distance. Specifically, it is a metric, it lies between 0 and 1 and it is equal to zero if and only if $F_1 \equiv F_2$. It is also obviously true that the Lévy distance is very sensitive to distribution F_1, F_2 changes.

A set of distances that are also very sensitive to distribution changes and are defined on Euclidean spaces are the Kolmogorov, variational, I-divergence and Bhattacharyya. The three last assume existence of non-cumulative distributions (density functions). If we concentrate on one-dimensional Euclidean spaces and denote by f_1, f_2 the densities of F_1, F_2 correspondingly, we denote:

Kolmogorov distance:

$$d_K(F_1, F_2) = \sup | F_1(x) - F_2(x) | \tag{5}$$

Variational distance:

$$d_{Vr}(f_1, f_2) = \int_{-\infty}^{\infty} | f_1(x) - f_2(x) | dx \tag{6}$$

I-divergence distance:

$$d_I(f_1, f_2) = \frac{1}{2} \int_{-\infty}^{\infty} \left[f_1(x) - f_2(x) \right] \log \frac{f_1(x)}{f_2(x)} dx \tag{7}$$

Bhattacharyya distance:

$$d_B(f_1, f_2) = \ln \left[\int_{-\infty}^{\infty} f_1^{\frac{1}{2}}(x) \; f_2^{\frac{1}{2}}(x) dx \right]^{-1} \tag{8}$$

The Bhattacharyya distance is well known to be an upper bound to the probability of erratic decision between f_1 and f_2. Also, the variational distance in addition to being over sensitive to density changes, it is also very demanding. A more detailed comparative evaluation of the six distances covered in this preliminary discussion is presented in the following section.

3. Comparative evaluation

The relationship between Prokhorov and Vasershtein distances has been already given by (3). The following lemma describes a simple relationship between the Lévy, Kolmogorov, and variational distances, where multidimensional distributions are in general considered.

Lemma 1

For Q_1, Q_2 multidimensional cumulative distributions and f_1, f_2 their corresponding densities, the following ranking is true:

$$d_L(Q_1,Q_2) \le d_K(Q_1,Q_2) \le d_{Vr}(f_1,f_2) \tag{9}$$

Proof:

1. For the relationship between Lévy and Kolmogorov distances, let

$$d_K(Q_1,Q_2)=\delta=\sup_X \left| Q_1(X)-Q_2(X) \right| .$$

Then,

$$Q_1(X)-Q_2(X)\le|Q_1(X)-Q_2(X)|\le\delta\le$$

$$\le Q_2(\cup Y:\rho(X,Y)\le\delta)-Q_2(X)+\delta,\forall X$$

and

$$Q_2(X)-Q_1(X)\le|Q_2(X)-Q_1(X)|\le\delta\le$$

$$\le Q_1(\cup Y:\rho(X,Y)\le\delta)-Q_1(X)+\delta;\forall X$$

or

$$\begin{cases} Q_1(X)\le Q_2(\cup Y:\rho(X,Y)\le\delta)+\delta;\forall X \\ Q_2(X)\le Q_1(\cup Y:\rho(X,Y)\le\delta)+\delta,\forall X \end{cases}$$

Therefore, δ is a candidate for $d_L(Q_1,Q_2)$ and $d_L(Q_1,Q_2)\le\delta=d_K(Q_1,Q_2)$.

2. For the relationship between variational and Kolmogorov distances:

$$|Q_1(X)-Q_2(X)|=\left|\int_{-\infty}^{X}\left[f_1(Y)-f_2(Y)\right]dY\right|\le$$

$$\le\int_{-\infty}^{X}|f_1(Y)-f_2(Y)|\,dY\le\int_{-\infty}^{\infty}|f_1(Y)-f_2(Y)|\,dY$$

$$\Rightarrow d_K(Q_1,Q_2)=\sup_X|Q_1(X)-Q_2(X)|\le d_{Vr}(f_1,f_2)$$

It can be seen from the above lemma that the Lévy distance is the weakest, the variational is the strongest and the Kolmogorov lies in between.

The Bhattacharyya and I-divergence distances consist a separate group that is not directly comparable to the previously mentioned distances. In addition, as explained by Kailath [8], the Bhattacharyya and I-divergence distances are not directly comparable either.

At this point it is interesting to make the peripheral comment that the nonsymmetric I-divergence distance given by the expression

$$d_I(f_1,f_2)=\int f_1(Y)\log\frac{f_1(Y)}{f_2(Y)}\,dY \quad (10)$$

is also the discrimination or relative entropy of f_2 with respect to f_1, defined by Wyner [19] for a two-receiver broadcast channel. The symmetric I-divergence distance in (7) can then be looked at as the mutual relative entropy between the two receivers of the same channel. Continuing on the comparative evaluations of the distances, we will present two lemmas involving the Prokhorov, generalized Kolmogorov and Vasershtein distances. The conclusion from these two lemmas is that for particular choices of penalty or distortion measures $\rho(.,.)$ both the Prokhorov and Vasershtein distances degenerate to the Kolmogorov one.

Lemma 2

Let $(\Omega,\mathcal{G},\mu_1)$ and $(\Omega,\mathcal{G},\mu_2)$ be two measure spaces with measures such that:

$$\mu_1(\Omega)=\mu_2(\Omega)=S,\ 0\le\mu_i(A)\le\mu_i(B);\forall A\subset B:A,B,\in\mathcal{G};i=1,2$$

where S some real number.

Let also $$\rho(A,B)\begin{cases} S, A\equiv B\\ 0, A\equiv B\end{cases}$$

Then $$d_{P,\rho}(\mu_1,\mu_2)=d_K(\mu_1,\mu_2)$$

<u>Lemma 3</u>

For measure spaces, distortion measure $\rho(.,.)$ as in lemma 2, and S=1, it is also true that

$$d_{V,\rho}(\mu_1,\mu_2)=d_K(\mu_1,\mu_2)$$

The result expressed by lemma 3 was first found by Dobrushin [13]. An alternative proof for this lemma and proof of lemma 2 are presented in appendix A.

Another property of the distances that is valuable for any optimization problem involving them is convexity. It is easy to show that the variational, Kolmogorov and I-divergence distances are convex $\cup$ on closed linear manifolds of any of the two distributions involved.

The Vasershtein distance for distributions and the Lévy distance are considered interesting cases here and are examined in detail.

Let us first consider the Vasershtein distance for distributions, and denote $Q_1(X_n),Q_2(Y_n)$ two n-dimensional distributions at X_n and Y_n respectively. In addition, denote by $R_{Q_1,Q_2}(X_n,Y_n)$ any 2n-dimensional distribution with Q_1,Q_2 marginals. For such distributions the Vasershtein distance is given by the following expression:

$$d_{V,\rho}(Q_1,Q_2)=\inf_{R_{Q_1;Q_2}(.,.)} E_{R_{Q_1;Q_q}(.,.)}\{\rho(X_n,Y_n)\} \qquad (11)$$

(11)

The following lemma can then be stated.

<u>Lemma 4</u>

For distortion measure $\rho(.,.)\geq 0$, the Vasershtein distance $d_{V,\rho}(Q_1,Q_2)$ in (11) is convex $\cup$ on any closed linear manifold of either Q_1, or Q_2.

Proof

Let $\mathcal{B}_{ij}$ be the space of all joint distributions R(.,.) that have Q_i, Q_j marginals. Consider the $\mathcal{B}_{12}, \mathcal{B}_{13}$ spaces and let $R_1(.,.) \in \mathcal{B}_{12}$, $R_2(.,.) \in \mathcal{B}_{13}$. Then, for every $h \in [0,1]$, the distribution

$$R(.,.) = hR_1(.,.) + (1-h)R_2(.,.)$$

has marginals Q_1 and $hQ_2+(1-h)Q_3$. Therefore, for every $R_1(.,.) \in \mathcal{B}_{12}$, $h \in [0,1]$, $R_2(.,.) \in \mathcal{B}_{13}$, the distribution

$$R(.,.)\ h(R_1(.,.) + (1-h)R_2(.,.) \tag{12}$$

is a member of the space $\mathcal{B}_{1h23}$ that has $Q_1, hQ_2+(1-h)Q_3$ marginals. In other words, the distributions expressed b (12) consist a subspace of the space $\mathcal{B}_{1h23}$.

Therefore,

$$d_{V,\rho}(Q_1, hQ_2+(1-h)Q_3) = \inf_{R(.,.)\in\mathcal{B}_{1h23}} E_{R(.,.)}\{\rho(X_n,Y_n)\} \leq$$

$$\leq \inf_{\substack{R(.,.)\text{ in the linear}\\ \text{family described by (12)}}} E_{R(.,.)}\{\rho(X_n,Y_n)\} = \inf_{\substack{R_1(.,.)\in\mathcal{B}_{12}\\ R_2(.,.)\in\mathcal{B}_{13}}} E_{hR_1(.,.)+(1-h)R_2(.,.)}\{\rho(X_n,Y_n)\}$$

$$= \inf_{\substack{R_1(.,.)\in\mathcal{B}_{12}\\ R_2(.,.)\in\mathcal{B}_{13}}} \left[hE_{R_1(.,.)}\{\rho(X_n,Y_n)\} + (1-h)E_{R_2(.,.)}\{\rho(X_n,Y_n)\}\right]$$

But since $\rho(.,.) \geq 0$, we can write from above:

$$d_{V,\rho}(Q_1, hQ_2+(1-h)Q_3) = h \inf_{R_1(.,.)\in\mathcal{B}_{12}} E_{R_1(.,.)}\{\rho(X_n,Y_n)\} +$$

$$+(1-h) \inf_{R_2(.,.)\in\mathcal{B}_{13}} E_{R_2(.,.)}\{\rho(X_n,Y_n)\}$$

$$= hd_{V,\rho}(Q_1,Q_2) + (1-h)d_{V,\rho}(Q_1,Q_3)$$

and the proof is here complete, since symmetric analysis leads to similar result for the other distributions involved in the Vasershtein distance.

We should make here the additional observation that the result of the above lemma can be extended to arbitrary measures μ_1,μ_2 and $\rho(.,.)\geq 0$.

Also, we will emphasize here that if the distortion measure $\rho(X_n,Y_n)$ is symmetric with respect to X_n and Y_n, the Vasershtein distance is also symmetric with respect to the distributions Q_1,Q_2. That is, $d_{V,\rho}(Q_1,Q_2)=d_{V,\rho}(Q_2,Q_1)$ then.

At the study of the Lévy distance for convexity, it became evident that such property is secured only if the distance is redefined on a closed internal of the distribution domain. The reason that such redefinition is necessary is that convexity of the underline distributions is required then, and such convexity is not true on the whole $(-\infty,\infty)$ for nontrivial such distributions. That will be clear in the following detailed discussion.

To make the discussion as meaningful as possible we will restrict ourselves to one-dimensional distributions F. The arguments and the results can be easily extended to multidimensional spaces.

Let us consider two constants a,b such that $a<b$ and define the Lévy distance in the following way:

$$d_{L,\rho}(F_1,F_2)=\inf\left\{\epsilon: F_2(\min(y:\rho(x,y)\leq\epsilon))-\epsilon\leq F_1(x)\leq F_2(\max(y:\rho(x,y)\leq\epsilon))+\epsilon;\ \forall x\in[a,b]\right\} \quad (13)$$

What is implied in the definition (13) is that the domain of interest is $[a,b]$, which means that the values of F_1,F_2 in the remaining domain are either kept fixed or they have no influence on the system under consideration.

Now, we can present the following lemma:

<u>Lemma 5</u>

The Lévy distance expressed by (13) is convex $\cup$ on the closed linear manifold of distributions F_1,F_2 that are convex $\cap$ on $[a-1,b+1]$.

Proof: _

We will prove the lemma for F_2 only. Due to symmetry the proof for F_1 will be the same.

Let F_1, F_{21}, F_{22} be three distributions that are $\cap$ convex on $[a,b]$, and form the distribution $F_2 = hF_{12} + (1-h)F_{22}$; where $0 \le h \le 1$. The distribution F_{123} is also convex $\cap$ on $[a,b]$. We can write as discussed in [21]:

$$d_{L,\rho}(F_1,F_{21}) = \inf\Big\{\varepsilon : F_{21}(x) - \Big[F_1(\max(y:\rho(x,y)\le\varepsilon)) + \varepsilon\Big] \le 0;$$
$$F_{21}(x) - \Big[F_1(\min(y:\rho(x,y)\le\varepsilon)) - \varepsilon\Big] \ge 0; \forall x \in \Big[0;b\Big]\Big\} = \varepsilon_1$$

$$d_{L,\rho}(F_1,F_{22}) = \inf\Big\{\varepsilon : F_{22}(x) - \Big[F_1(\max(y:\rho(x,y)\le\varepsilon)) + \varepsilon\Big] \le 0;$$
$$F_{22}(x) - \Big[F_1(\min(y:\rho(x,y)\le\varepsilon)) - \varepsilon\Big] \ge 0; \forall x \in \Big[a,b\Big]\Big\} = \varepsilon_2 \tag{14}$$

$$d_{L,\rho}(F_1,F_2) = \inf\Big\{\varepsilon : h\Big[F_{21}(x) - [F_1(\max(y:\rho(x,y)\le\varepsilon)) + \varepsilon]\Big] + (1-h)\Big[F_{22}(x) -$$
$$-\Big[F_1(\max(y:\rho(x,y)\le\varepsilon)) + \varepsilon\Big)\Big] \le 0, h\Big[F_{22}(x) - \Big[F_1(\min(y:\rho(x,y)\le\varepsilon)) - \varepsilon\Big]\Big] +$$
$$+(1-h)\Big[F_{22}(x) - \Big[(\min(y:\rho(x,y)\le\varepsilon) - \varepsilon\Big]\Big] \ge 0; \forall x \in \Big[a,b\Big]\Big\} \tag{15}$$

Since $d_{L,\rho}(F_1,F_{21}) = \varepsilon_1$, for every $\delta > 0$ there is some $\varepsilon_{\delta_1} : \varepsilon_1 \le \varepsilon_{\delta_1} < \varepsilon_1 + \delta$ that satisfies:

$$\begin{cases} F_{21}(x) - \Big[F_1(\max(y:\rho(x,y)\le\varepsilon_{\delta_1})) + \varepsilon_{\delta_1}\Big] \le 0 \\ F_{21}(x) - \Big[F_1(\min(y:\rho(x,y)\le\varepsilon_{\delta_1})) - \varepsilon_{\delta_1}\Big] \ge 0 \end{cases} \quad \forall x \in [a,b] \tag{16}$$

Similarly, from $d_{L,\rho}(F_1,F_{22})$ we obtain that for every $\delta > 0$ there is some

$\epsilon_{\delta_2}: \epsilon_2 \le \epsilon_{\delta_2} < \epsilon_2 + \delta$ that satisfies:

$$\begin{cases} F_{22}(x) - \left[F_1(\max(y:\rho(x,y)\le\epsilon_{\delta_2})) + \epsilon_{\delta_2}\right] \le 0 \\ \\ F_{22}(x) - \left[F_1(\min(y:\rho(x,y)\le\epsilon_{\delta_2})) - \epsilon_{\delta_2}\right] \ge 0 \end{cases} \quad \forall x \in [a,b] \tag{17}$$

From (16) and (17) it is implied that:

$$h\left[F_{21}(x) - \left[F_1(\max(y:\rho(x,y)\le\epsilon_{\delta_1})) + \epsilon_{\delta_1}\right]\right] + (1-h)\left[F_{22}(x) - \left[F_1(\max(y:\rho(x,y)\le\epsilon_{\delta_2})) + \epsilon_{\delta_2}\right]\right] \le 0$$

$$\forall x [a,b] \tag{18}$$

$$h\left[F_{21}(x) - \left[F_1(\min(y:\rho(x,y)\le\epsilon_{\delta_1})) - \epsilon_{\delta_1}\right]\right] + (1-h)\left[F_{22}(x) - \left[F_1(\min(y:\rho(x,y)\le\epsilon_{\delta_2})) - \epsilon_{\delta_2}\right]\right] \ge 0$$

or that:

$$hF_{21}(x) + (1-h)F_{22}(x) \le hF_1(\max(y:\rho(x,y)\le\epsilon_{\delta_1})) + (1-h)F_1(\max(y:\rho(x,y)\le\epsilon_{\delta_2})) + \left[h\epsilon_{\delta_1} + (1-h)\epsilon_{\delta_2}\right]$$

$$\forall x \in [a,b] \tag{19}$$

$$hF_{21}(x) + (1-h)F_{22}(x) \ge hF_1(\min(y:\rho(x,y)\le\epsilon_{\delta_1})) + (1-h)F_1(\min(y:\rho(x,y)\le\epsilon_{\delta_1})) - \left[h\epsilon_{\delta_1} + (1-h)\epsilon_{\delta_2}\right]$$

Due to the $\cap$ convexity of F_1 on $[a-1,b+1]$ we have:

$$F_1(h\max(y:\rho(x,y)\le\epsilon_{\delta_1}) + (1-h)\max(y:\rho(x,y)\le\epsilon_{\delta_2})) = F_1(x + h\rho^{-1}(\epsilon_{\delta_1}) + (1-h)\rho^{-1}(\epsilon_{\delta_2}))$$

$$\ge hF_1(x+\rho^{-1}(\epsilon_{\delta_1})) + (1-h)F_1(x+\rho^{-1}(\epsilon_{\delta_2})) \tag{20}$$

where $\rho^{-1}(\epsilon) = \max(y:\rho(x,y)\le\epsilon) - x$

Substituting (20) in the first part of (19) we obtain:

$$hF_{21}(x) + (1-h)F_{22}(x) \le F_1(x + h\rho^{-1}(\epsilon_{\delta_1}) + (1-h)\rho^{-1}(\epsilon_{\delta_2})) + h\epsilon_{\delta_1} + (1-h)\epsilon_{\delta_2}$$

$$\forall x \in [a,b] \tag{21}$$

The second part of (19) can be treated the same way as dicussed in [21]. If $\rho^{-1}(\epsilon)\leq\epsilon$ it is directly then derived from (21) that

$$d_{L,\rho}(F_1,hF_{21}+(1-h)F_{22})\leq h\epsilon_{\delta_1}+(1-h)\epsilon_{\delta_2}<\delta+h\epsilon_1+(1-h)\epsilon_2;\forall\delta>0,\rho(x,y)=|x-y|$$

Therefore, $d_{L,\rho}(F_1,hF_{21}+(1-h)F_{22})\leq hd_{L,\rho}(F_1,F_{21})+(1-h)d_{L,\rho}(F_1,F_{22})$ and the proof is complete

---

We can observe here that a distribution F(x) that is convex $\cap$ on $[a-1,b+1]$ is a distribution with negative second derivative in $[a-1,b+1]$ which corresponds to density monotomically decreasing on the same interval.

As a conclusion from lemma 5 we can observe that due to the convexity expressed there, given any distribution F there is a unique best approximation of it in the Lévy distance sense given by (12), where the approximation is taken from the linear manifold of distributions that are convex $\cap$ inside a certain closed interval $[a-1,b+1]$.

As we see in the following section, the convexity of the distances besides been valuable for best approximation problems, it is also useful in the search for optimal linear data transformations under restrictions.

4. The distances and data linear transformations

Let $(\Omega,\mathcal{G})$ be a space with a σ-algebra defined on it, and let T be a transformation on the ω elements of the space Ω. The transformed space will be called $T\Omega$, and the transformed σ-algebra will be denoted $T\mathcal{G}$.

Corollary 1

$T\mathcal{G}$ is a σ-algebra if Ω is a countable space.

Proof

Let $A,B\in\mathcal{G}$ and denote by TA, TB the transformed sets. Then $TA\cup TB$ and $TA\cap TB$ are obviously equal to $TA\cup B$ and $TA\cap B$ correspondingly. But, $A,B,A\cup B,A\cap B\in\mathcal{G}$. Therefore, $TA,TB,TA\cup B,TA\cap B\in T\mathcal{G}$. The same is easily extended to $\bigcup_i TA_i$ and

$$\lim_{n\to\infty} \bigcup_{i=1}^{n} TA_i .$$

We want to point out here that if Ω is not countable the statement of the corollary is not necessarily true. Also, if Ω is countable so is $T\Omega$, while T is not necessarily a one-to-one transformation.

Now, suppose that a measure space $(\Omega,\mathcal{G},\mu)$ and a closed convex family $\mathcal{F}$ of transformations T are given, such that for every $T\in\mathcal{F}$, $T\mathcal{G}$ is a σ-algebra. Then, each $T\in\mathcal{F}$ induces a unique measure space $(T\Omega,T\mathcal{G},\nu_T(\mu))$.

Let two different measures μ_1,μ_2 assigned as $(\Omega,\mathcal{G})$ and let T be some transformation from the family $\mathcal{F}$. Then, two measures $\nu_T(\mu_1)$ and $\nu_T(\mu_2)$ defined on $(T\Omega,T\mathcal{G})$ are induced by T,μ_1,μ_2. If some distance $d(\nu_T(\mu_1),\nu_T(\mu_2))$ is convex $\cap$ with respect to T for $T\in\mathcal{F}$, then there is a unique transformation $T_o\in\mathcal{F}$ that applied on the measure spaces $(\Omega,\mathcal{G},\mu_1)$, $(\Omega,\mathcal{G},\mu_2)$ induces measures $\nu_{T_o}(\mu_1)$, $\nu_{T_o}(\mu_2)$ that realize the $\max_{T\in\mathcal{F}} d(\nu_T(\mu_1),\nu_T(\mu_2))$.

In this section we will be concerned with a particular space $(\Omega,\mathcal{G})$, and particular measures μ_1,μ_2 and transformation T. Specifically, Ω will be the E^n Euclidean space and the σ-algebra $\mathcal{G}$ will include all sets $(-\infty,X_n)$ where X_n an n-dimensional vector defined on E^n. The family $\mathcal{G}$ of transformations will be the family of row vectors C_n of dimensionality n that satisfy some restriction. If this restriction is $C_n R C_n' \leq \alpha$, where α some constant, and R same nxn nomegative matrix, the family of transformations is convex. The measures μ_1,μ_2 will be, in general, pro-

bability measures, and more specifically Gaussian probability measures defined on $(E^n, \mathcal{G})$.

Let

$$\mu_i(-\infty, X_n) = \int_{-\infty}^{X_n} f_{ni}(Y_n)\,dY_n ; i=1,2 \tag{22}$$

where

$$f_{ni}(Y_n) = (2\pi)^{-\frac{n}{2}} \left| R_{ni} \right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(Y_n - M_{ni})' R_{ni}^{-1} (Y_n - M_{ni}) \right\} ; i=1,2 \tag{23}$$

Then, for $T=C_n$ we obtain from (22),(23):

$$\nu_T(\mu_i)(-\infty, Z) = \int_{-\infty}^{Z} f_{iT}(w)\,dw ; i=1,2 \tag{24}$$

where

$$f_{iT}(w) = (2\pi)^{-\frac{1}{2}} \left[C_n R_{ni} C_n' \right]^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \left[\frac{(w - C_n M_{ni})^2}{C_n R_{ni} C_n'} \right] \right\} ; \ i=1,2 \tag{25}$$

Finally, let the convex family of transformations $\mathcal{T}$ be expressed by the restriction

$$C_n R_{n1} C_n' \leq A \tag{26}$$

where R_{n1} the positive definite covariance matrix in (23) corresponding to i=1.

The optimal transformation $\overset{\circ}{C}_n$ (if any) that has the best discriminant effects will be sought, where this effect is measured by the Bhattacharrya, I-divergence, Lévy, Vasershtein, Kolmogorov and variational distances. Each of the distances will be examined separately, for the two cases of $M_{n1}=M_{n2}$, $R_{n1}=R_{n2}$.

a. $\underline{M_{n1}=M_{n2}=M_n}$

I. <u>Bhattacharyya distance</u>

In this case, the Bhattacharyya distance in (8) becomes:

$$d_B(f_{1T}, f_{2T}) = \ln \sqrt{ \frac{1}{2} \left(\frac{C_n R_{n1} C_n'}{C_n R_{n2} C_n'} \right)^{\frac{1}{2}} + \frac{1}{2} \left(\frac{C_n R_{n2} C_n'}{C_n R_{n1} C_n'} \right)^{\frac{1}{2}} } \tag{27}$$

Let $R_{n1}=WW'$; where W the matrix of the relative eigenvectors and $R_{n2}=WLW'$; where L the diagonal matrix with the eigenvalues λ_j of R_{n2} w.r.t. R_{n1}. Then, the optimal transformation C_n° that maximizes $d_B(f_{1T},f_{2T})$ under the restriction in (26) is actually a set of infinite values lying on a line and described by:

$$C_n^{\circ} = K\max^{-1}\left(\lambda_{max}^{\frac{1}{2}}+\lambda_{max}^{-\frac{1}{2}},\lambda_{min}^{\frac{1}{2}}+\lambda_{min}^{-\frac{1}{2}}\right) \tag{28}$$

where K any constant that does not exceed absolutely $\sqrt{A}$.

$\max^{-1}\left(\lambda_{max}^{\frac{1}{2}}+\lambda_{max}^{-\frac{1}{2}},\lambda_{min}^{\frac{1}{2}}+\lambda_{min}^{-\frac{1}{2}}\right)$ = this C_n from $C_n^{\lambda_{min}}, C_n^{\lambda_{max}}$ that realizes the maximum between $\lambda_{max}^{\frac{1}{2}}+\lambda_{max}^{-\frac{1}{2}}$, $\lambda_{min}^{\frac{1}{2}}+\lambda_{min}^{-\frac{1}{2}}$

and
$$C_n^{\lambda_{min}} = [0,..010..0]W^{-1} \tag{29}$$

↑ position corresponding to the minimum L eigenvalue

$$C_n^{\lambda_{max}} = [0,..010..0]W^{-1} \tag{30}$$

↑ position corresponding to the maximum eigenvalue in L

The proof of this result is in appendix B.

The maximum distance is given by the following expression:

$$d_{Bmax}=-\frac{1}{2}\ell n2+\frac{1}{2}\ell n\max\left(\lambda_{max}^{\frac{1}{2}}+\lambda_{max}^{-\frac{1}{2}},\lambda_{min}^{\frac{1}{2}}+\lambda_{min}^{-\frac{1}{2}}\right) \tag{31}$$

II. <u>Variational distance</u>

$$d_{V_r}\left(f_{1T},f_{2T}\right)=4\left|\Phi\left(\sqrt{\frac{2\ell n\frac{C_nR_{n2}C_n'}{C_nR_{n1}C_n'}}{\left(\frac{C_nR_{n2}C_n'}{C_nR_{n1}C_n'}\right)^2-1}}\right)-\Phi\left(\frac{C_nR_{n2}C_n'}{C_nR_{n1}C_n'}\sqrt{\frac{2\ell n\frac{C_nR_{n2}C_n'}{C_nR_{n1}C'}}{\left(\frac{C_nR_{n2}C_n'}{C_nR_{n1}C_n'}\right)^2-1}}\right)\right| \tag{32}$$

where $\Phi(x)=\int_{-\infty}^{x} \frac{1}{2\pi} \exp\left\{-\frac{u^2}{2}\right\} du$

as shown in appendix B. Again the maximum distance corresponds to

$$\frac{C_n R_{n2} C_n'}{C_n R_{n1} C_n'} = \max(\lambda_{max}, \lambda_{min}^{-1}) \tag{33}$$

where $\lambda_{max}, \lambda_{min}$ the maximum and minimum eigenvalues in L correspondingly. The corresponding linear "optimal" transformations are again given by (29) and (30) multiplied by any constant not exceeding $\sqrt{A}$.

The maximum value of the distance is:

$$d_{Vrmax} = 4 \max\left(\left| \Phi\left(\sqrt{\frac{2\ln\lambda_{max}}{\lambda_{max}^2 - 1}} \right) - \Phi\left(\lambda_{max} \sqrt{\frac{2\ln\lambda_{max}}{\lambda_{max}^2 - 1}} \right) \right| , \left| \Phi\left(\sqrt{\frac{2\ln\lambda_{min}}{\lambda_{min}^2 - 1}} \right) - \Phi\left(\lambda_{min} \sqrt{\frac{2\ln\lambda_{min}}{\lambda_{min}^2 - 1}} \right) \right| \right) \tag{34}$$

III. Kolmogorov distance

It can be found easily that

$$d_K(f_{1T}, f_{2T}) = \left| \Phi\left(\frac{C_n R_{n2} C_n'}{C_n R_{n1} C_n'} \sqrt{\frac{2\ln \frac{C_n R_{n2} C_n'}{C_n R_{n1} C_n'}}{\left(\frac{C_n R_{n2} C_n'}{C_n R_{n1} C_n'}\right)^2 - 1}} \right) - \Phi\left(\sqrt{\frac{2\ln \frac{C_n R_{n2} C_n'}{C_n R_{n1} C_n'}}{\left(\frac{C_n R_{n2} C_n'}{C_n R_{n1} C_n'}\right)^2 - 1}} \right) \right| \tag{35}$$

The same transformations $C_n^{\lambda}{}_{min}$, $C_n^{\lambda}{}_{max}$ in (29), (30), (28) stand and

$$d_{Kmax} = \frac{1}{4} d_{Vrmax} \tag{36}$$

where d_{Vrmax} is given by (34)

IV. I-divergence distance

It can easily be found that:

$$d_I\left(f_{1T},f_{2T}\right) = -\frac{1}{2} + \frac{1}{4}\left[\frac{C_nR_{n2}C_n'}{C_nR_{n1}C_n'} + \frac{C_nR_{n1}C_n'}{C_nR_{n2}C_n'}\right] \tag{37}$$

It is obvious that in (37) the same symmetries that appear in (27) exist.

Therefore, the "optimal" linear transormation is the same with the one in the Bhattacharyya distance. The maximum distance value is given by:

$$d_{Imax} = -\frac{1}{2} + \frac{1}{4}\max\left(\lambda_{max}+\lambda_{max}^{-1},\ \lambda_{min}+\lambda_{min}^{-1}\right) \tag{38}$$

V. Lévy distance

We felt the Lévy and Vasershtein distances last because they both involve a penalty or distortion measure and therefore consist a separate group.

If $\rho(.,.)$ is the distortion measure used; denote by $\rho^{-1}(\varepsilon)$ the maximum y that gives $\rho(x,x+y)=\varepsilon$ for any x. Then, the Lévy distance for the equal mean model considered here is given by:

$$d_{L,\rho}\left(f_{1T},f_{2T}\right) = \inf\left\{\varepsilon : \Phi(y)-\Phi\left(y\frac{C_nR_{n1}C_n'}{C_nR_{n2}C_n'} + \frac{\rho^{-1}(\varepsilon)}{C_nR_{n2}C_n'}\right) \le \varepsilon;\ \Phi\left(y\frac{C_nR_{n1}C_n'}{C_nR_{n2}C_n'}\right)-\Phi\left(y+\frac{\rho^{-1}(\varepsilon)}{C_nR_{n1}C_n'}\right) \le \varepsilon; \forall_y\right\} \tag{39}$$

As shown in appendix B, finding the Lévy distance for given C_n transformations or even more finding the "optimal" discriminant transformation in this case becomes a task that can be only approached numerically. This is a strong indication of the fact that the Prokhorov and Lévy distances are characterized by the disadvantage of difficult calculability. In cases that the inclusion of a penalty or distortion measure is desirable, the Vasershtein distance is a better choice as we will show below.

VI. Vasershtein distance

We will work with specific distortion measures here. The first such distortion measure to be examined will be the popular $\rho_s(x,y)=(x-y)^2$.

For the equal mean case we are considering here, it can be shown directly:

$$d_{V,\rho}(f_{1T},f_{2T})=\inf_{\substack{r(.,.)\text{with}\\ f_{1T},f_{2T}\text{ marginals}}} \left\{C_n R_{n1} C_n'+C_n R_{n2} C_n'-2E_{r(.,.)}\left\{(y-C_n M_n)(x-C_n M_n)\right\}\right\} \tag{40}$$

As shown in [16] the Vasershtein distance $d_{V,\rho_s}(f_{1T},f_{2T})$ can be actually found if the knowledge of some structure involving the intial data X_n, Y_n is assumed.

Specifically, let X_n be n samples from or Gaussian, wide sense stationary process whose nth order statistics are described by $f_{n1}(X_n)$ in (23). Then, the autocorrelations matrix R_{n1} is a Toeplitz matrix. Let Y_n be from a Gaussian wide sense stationary process also whose nth order statistics are described by $f_{n2}(Y_n)$ in (23) and R_{n2} is again Toeplitz. Then the scalar variables $C_n X_n, C_n Y_n$ are samples from Gaussian, wide sense stationary processes also. In fact, the power spectra of the variables $C_n X_n, C_n Y_n$ exist and are given by the following expressions correspondingly.

$$\mathcal{P}_{f_{1T}}(\lambda) = \sum_{k=-\infty}^{\infty} C_n R_{n1}(k) C_n' \, e^{ik\lambda} \tag{41}$$

$$\mathcal{P}_{f_{2T}}(\lambda) = \sum_{k=-\infty}^{\infty} C_n R_{n2}(k) C_n' \, e^{ik\lambda} \tag{42}$$

where $\mathcal{P}_{f_{iT}}(\lambda)$; i=1,2 denotes power spectrum under corresponding n order statistics in (23), and C_n transformation. Also, $R_{ni}(k)=E\left\{\left(X_n-M_{ni}\right)\left(X_{n+k}-M_{n+k,i}\right)'\right\}$, where the expectation is taken over the statistics that in n order are given by f_{ni} in (23), and what has been denoted by R_{ni} till now is actually $R_{ni}(o)$.

Let us restrict ourselves to cross-stationary processes. Then the infimum in (40) will be taken over all cross-stationary statistics with f_{1T},f_{2T} marginals, and the cross power spectrum exists and is denoted by $\mathcal{P}_{f_{1T},f_{2T}}(\lambda)$.

As explained in [16] p. 324, the Vasershtein distance under the above restrictions is given by the following expression:

$$d_{V,\rho_s}(f_{1T},f_{2T})=(2\pi)^{-1}\int_{-\pi}^{\pi}\left|\rho^{\frac{1}{2}}_{f_{1T}}(\lambda)-\rho^{\frac{1}{2}}_{f_{2T}}(\lambda)\right|^2 d\lambda \tag{43}$$

where $\rho_{iT}(\lambda)$ are given by (41) and (42). We want to point out here that for the existence of $\rho_{iT}(\lambda)$ it is sufficient that the vectors $X_n Y_n$ form cross stationary processes in time. In this case, the matrices $R_{n1}(0), R_{n2}(0)$ do not have to be Toeplitz.

Let $X_n(j)$, $Y_n(j)$ denote n-dimensional data, collected at time j from populations distributed as in f_{n1}, f_{n2} correspondingly. Let both $X_n(j), Y_n(j)$; $j=0,1,...$ be first order Markov. Then, the spectra in (41) and (42) become:

$$\rho_{f_{1T}}(\lambda)=C_n R_{n1}(o)C_n'+2C_n R_{n1}(1)C_n'\cdot\cos\lambda \tag{44}$$

$$\rho_{f_{2T}}(\lambda)=C_n R_{n2}(o)C_n'+2C_n R_{n2}(1)C_n'\cdot\cos\lambda \tag{45}$$

If expressions (44) and (45) are substituted in (43) the following expression is obtained:

$$d_{V,\rho_s}(f_{1T},f_{2T})=C_n[R_{n1}(o)+R_{n2}(o)]C_n' -$$

$$-\pi^{-1}\int^{\pi}\left\{C_n\left[R_{n1}(o)+2\cos\lambda R_{n1}(1)\right]C_n'\cdot C_n\left[R_{n2}(o)+2\cos\lambda R_{n2}(1)\right]C_n'\right\}^{\frac{1}{2}}d\lambda \tag{46}$$

Calculating the expression in (46) analytically to study extremes with respect to the linear transformation C_n is not possible. For that reason we are using bounds. Indeed, applying the Schwartz inequality on the integral in (46) we obtain:

$$d_{V,\rho_s}(f_{1T},f_{2T})\geq\left[\sqrt{C_n R_{n1}(o)C_n'} - \sqrt{C_n R_{n2}(o)C_n'}\right]^2 \tag{47}$$

with equality if and only if there is some constant B such that:

$$C_n\left(\left[R_{n1}(0)-BR_{n2}(0)\right]+2\cos\lambda\left[R_{n1}(1)-BR_{n2}(1)\right]\right)C_n' = 0 \tag{48}$$

for almost all $\lambda \in[-\pi,\pi]$.

To maximize $d_{V,\rho_s}(f_{1T},f_{2T})$ with respect to the C_n choice we will maximize the lower bound in (47) instead. Since this bound can be written:

$$B(\rho_s)=C_nR_{n1}(o)C_n' \left[1 - \sqrt{\frac{C_nR_{n2}(o)C_n'}{C_nR_{n1}(o)C_n'}} \right]^2 \tag{49}$$

and due to the fact that, as shown in appendix B, the ratio $\frac{C_nR_{n2}(o)C_n'}{C_nR_{n1}(o)C_n'}$ can only vary in $\left[\lambda_{min},\lambda_{max}\right]$ with C_n changing (where λ_i the eigenvalues in L), the bound in (49) can increase to infinity for unrestricted C_n transformations. However, if (26) is true the bound in (49) is maximized for:

$$C_n=\sqrt{A}\ [0\ldots010..0]W^{-1} \tag{50}$$

where the value 1 in [0...010..0] corresponds to this position in L that belongs to max $(\lambda_{max},\lambda^{-1}min)$. The value of this maximum bound is given by:

$$B(\rho_s)=A\left[1- \max(\lambda_{max},\lambda^{-1}min)\right]^2 \tag{51}$$

We want to emphasize here that the Vasershtein distance as well as the bound in (49) depend only on the second order statistics of the data $X_n(j),Y_n(j)$ and they are totally insensitive to the exact underline distribution. This property is valuable in the case of ill-defined environments.

If the distortion measure is $\rho_\ell(x,y)=|x-y|$ instead of $(x-y)^2$, the Vasershtein distance $d_{V,\rho_\ell}(f_{1T},f_{2T})$ is bounded from below ([16], th. 5) by

$$B(\rho_\ell)=(2\pi)^{\frac{1}{2}}[C_nR_{n1}(0)C_n'-C_nR_{n2}(0)C_n'] \tag{52}$$

The bound in (52) is obviously maximized for the transformation in (50), where the restriction (26) is again true. This maximum is given by the following expression:

$$B_{max}(\rho_\ell) = 2\pi^{\frac{1}{2}} A\left[1-\max\left(\lambda_{max}, \lambda^{-1}_{min}\right)\right] \tag{53}$$

VII. General Observations

It is evident from the preceding analysis that for data X_n, Y_n distributed as described by $f_{n1}(X_n), f_{n2}(Y_n)$ in (23) correspondingly with $M_{n1}=M_{n2}$, and restricted linear transformations C_n, where the restriction is described by (26), the Bhattacharyya, I-divergence, Kolomogorov and variational distances are all maximized by the same transformation

$$C_n = K[0\ldots010\ldots0]W^{-1} \tag{54}$$

where $R_{n1}=WW'$, $R_{n2}=WLW'$, $L=\{\lambda_i\}$ and the digit 1 in (54) corresponds to the position of max $(\lambda_{max}, \lambda^{-1}_{min})$. The amplitude K is such that: $K\in[-\sqrt{A}, \sqrt{A}\ ]$.

The *Vasershtein distance with the implication of* $X_n(j), Y_n(j)$; $j=0,1\ldots$ being n dimensional stationary, first order Markov processes and with underline distortion measures either $\rho_s(x,y)=(x-y)^2$ or $\rho_\ell(x,y)=|x-y|$, does not require the specific statistics described by (23). In addition, a certain lower bound on this distance is maximized again by the transformations C_n in (54), where the amplitude K is equal to $\sqrt{A}$.

b. $R_{n1}=R_{n2}=R_n$; $M_n = M_{n2}-M_{n1}$

1. Bhattacharyya distance

The Bhattacharyya distance is given in this case by the following expression:

$$d_B(f_{1T}, f_{2T}) = \frac{1}{8}\,\frac{(C_n M_n)^2}{C_n R_n C_n'} \tag{55}$$

Our objective is to maximize the ratio in (55) with respect to linear transformations C_n that satisfy the restriction in (26).

It is straight forward to obtain the following expression for the "optimal" C_n:

$$C_n = K M_n' R_n^{-1} \quad ; \quad K \in [-\sqrt{A}, \sqrt{A}] \tag{56}$$

The maximum value of the distance is:

$$d_B = \frac{1}{8} M_n' R_n^{-1} M_n \tag{57}$$

I.Variational distance

It is easy to find that here:

$$d_{Vr}(f_{1T}, f_{2T}) = 2 \left| \Phi\left(\frac{C_n M_n}{2\sqrt{C_n R_n C_n'}}\right) - \Phi\left(-\frac{C_n M_n}{2\sqrt{C_n R_n C_n'}}\right) \right| \tag{58}$$

where,

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{u^2}{2}\} du$$

The distance in (58) is maximized for this C_n that realizes the maximum $\left| \frac{C_n M_n}{2\sqrt{C_n R_n C_n'}} \right|$ which is the one in (55) again.

The maximum distance value is given by:

$$d_{Vrmax} = 2[\Phi(2^{-1}\sqrt{M_n' R_n^{-1} M_n}) - \Phi(-2^{-1}\sqrt{M_n' R_n^{-1} M_n})] \tag{59}$$

I . Kolmogorov distance

Then,

$$d_K(f_{1T}, f_{2T}) = \left| \Phi\left(\frac{C_n M_n}{2\sqrt{C_n R_n C_n'}}\right) - \Phi\left(\frac{C_n M_n}{2\sqrt{C_n R_n C_n'}}\right) \right| \tag{60}$$

$$d_{Kmax} = 2^{-1} d_{Vrmax} \tag{61}$$

where d_{Vrmax} is given by (59) and the "optimal" C_n is again the one in (56).

IV .I-divergence distance

It can be easily found again that in the present case:

$$d_I(f_{1T}, f_{2T}) = 2^{-1} \frac{(C_n M_n)^2}{C_n R_n C_n'} \tag{62}$$

and of course the maximum is also obtained by the C_n in (56) and

$$d_{1\,max} = \tfrac{1}{2} M_n' R_n^{-1} M_n \tag{63}$$

V. Lévy Distance

As we will see, in this case of equal convariances the "optical" transformation in the Lévy distance sense can be found. It is shown in appendix B that in the present case:

$$d_{L,\rho}(f_{1T},f_{2T}) = \inf \{\epsilon : \Phi\left(-\frac{\rho^{-1}(\epsilon)}{2\sqrt{C_n R_n C_n'}}\right) - \Phi\left(-\frac{C_n M_n}{2\sqrt{C_n R_n C_n'}}\right) \leq \epsilon\} \tag{64}$$

where $\rho^{-1}(\epsilon) = \max\,(y:(x,x+y) = \epsilon)$

For momentarily fixed power $C_n R_n C_n'$, the transformation that obtains maximum $d_{L,\rho}$ in (64) is the one that maximizes $C_n M_n$. That is because for $C_n M_n$ value equal to $x_1 < x_2$ it is obviously true that every ϵ candidate for $d_{L,\rho}(f_{1T},f_{2T})$ with $C_n M_n = x_2$ instead.

So, letting now $C_n R_n C_n'$ vary in $[o,A]$, we obtain the maximum Lévy distance in (64) for C_n as in (56) with K amplitude fixed and equal to $\sqrt{A}$.

VI. Vasershtein Distance

Let again (as in a_{VI}) $X_n(j)$, $Y_n(j)$; $j = o,\ldots$ be stationary n-dimensional processes that are also first order Markov. Let $X_n(j)$ come from a population with M_{n1} and $R_n(o)$, $R_n(1)$ and let $Y_n(j)$ come from another population with M_{n2} and same $R_n(o)$, $R_n(1)$. The spectra $P_{f_{1T}}(\lambda)$, $P_{f_{1T}}(\lambda)$ in (42), (43) are then equal and the Vasershtein distance for cross-stationary $X_n(j)$, $Y_n(j)$ is realized for joint spectrum

$P_{f_{1T}f_{2T}}(\lambda) = [\, P_{f_{1T}}(\lambda)\, P_{f_{2T}}(\lambda)\,]^{\frac{1}{2}}$ and it is equal to:

$$d_{V\rho_s}(f_{1T},\, f_{2T}) = (C_n M_n)^2 \quad * \tag{65}$$

where $\rho_s(x,y) = (x - y)^2$.

The distance in (65) is obviously maximized for C_n given by (56) with fixed amplitude $K = \sqrt{A}$.

The maximum value of the distance is:

$$d_{V\rho_s \max} = A(M_n R_n^{-1} M_n)^2 \tag{66}$$

If the distortion measure is $\rho_\ell(x,y) = |\, x - y \,|$, then

$$E\, \{\rho_\ell(x,y)\} = E\, \{ |(x-C_n M_{n1})-(y-C_n M_{n2})-C_n M_n| \} \tag{67}$$

where the expectation is over all joint statistics with f_{1T}, f_{2T} marginals. For this case of equal spectra, and due to the lower bound given by theorem 5 in [16] we have from (67):

$$\begin{aligned} E\, \{\rho_\ell(x,y)\} &\geq |C_n M_n| - E\, \{ |(x-C_n M_{n1}) - (y-C_n M_{n2})| \} \geq \\ &\geq |C_n M_n| - \pi^{-1} \left[\int_{-\pi}^{\pi} |P^{\frac{1}{2}}_{f_{1T}}(\lambda) - P^{\frac{1}{2}}_{f_{2T}}(\lambda)\, |\, d_\lambda \right]^{\frac{1}{2}} = \\ &= |\, C_n M_n | \end{aligned}$$

---

*Foot Note: For the structure considered here and in general $R_{n1}(k) \neq R_{n2}(k)$; $k = 0,1,\ldots$ $M_{n1} \neq M_{n2}$ the Vasershtein distance is given by the expression:

$$d_{V\rho_s}(f_{1T}, f_{2T}) = (2\pi)^{-1} \int_{-\pi}^{\pi} |\, P^{\frac{1}{2}}_{f_{1T}}(\lambda) - P^{\frac{1}{2}}_{f_{2T}}(\lambda)\, |\, d_\lambda + (C_n M_n)^2 \; .$$

and the rate distortion theory, have been evaluated and used for the linear reduction of Gaussian data to one scalar parameter. It was found that while the Bhattachayya, I-divergence, vatiational, kalmogorov and Lévy distances are over-sensitive to the underline statistics, the Vasershtein distance depends only on second order moments.

Also, while the Lévy distance is hard to calculate analytically even in the simple case of Gaussian data, simple lower bound on the Vasershtein distance can be found for the distortion measures $\rho_s(x,y) = (x-y)^2$ $\rho_\ell(x,y) = |x-y|$.

Finally, for the Gaussian data and the linear reduction mentioned above, it was found that all distances (whenever the result analytically feasible) give the same "optimal" transformation with the most highly class-discrimination properties. This is true in both equal mean and equal covariances cases.

REFERENCES

1. A. Bhattacharyya, (1943) "On a measure of divergence between two statistical populations defined by their probability distributions," Bull. Calcutta Math. Soc., vol. 35, pp. 99-109.

2. S. Kakutani, (1948) "On equivalence of infinite product measures," Am. Math. Stat., vol. 49, pp. 214-224.

3. S. Kullback and R.A. Leibler, (1951) "On information and sufficiency," Am. Math. Stat., vol. 22, pp. 79-86.

4. Yu. V. Prokhorov (1956),"Convergence of random processes and limit theorems in probability theory," Theor. Prob. Appl., vol. 1, pp. 157-214.

5. S. Kullback, (1959), Information Theory and Statistics, New York: Wiley.

6. C. Stein, (1964) "Approximation of improper prior measures by prior probability measures," Department of Statistics, Stanford University, Stanford, California Tech. Rept. 12.

7. V. Strassen (1965) "The existence of probability measures with given marginals," Am. Math. Statist. vol 36, pp. 423-439.

8. T. Kailath (1967), "The Divergence and Bhattacharyya Distance Measures in Signal Selection," IEEE Trans., Comm. Tech. vol. Comm. 15, No. 1, pp. 52-60.

9. S. Kullback (1967), "A lower bound for discrimination information in terms of variation," IEEE Trans. Infom. The. Vol. IT-13, pp. 126-127.

10. K.R Parthasarathy, (1967), Probability Measures on Metric Spaces, New York: Academic Press.

11. R.M. Dudley, (1968), "Distances of Probability Measures and Random Variables," Am. Math. Statist., vol. 39, No. 5, pp. 1563-1572.

12. L.N. Vasershtein (1969), "Markov processes on a countable product space, describing large systems of automata," Problemy Peredachi Informatsii, vol. 5,3, pp. 64-73 (in Russian).

13. R.L. Dobrushin, (1970) "Prescribing a System of Random Variables by Conditional Distributions," Theory of Prob. and Appl., vol. XV, No. 3, pp. 458-486.

14. L.N. Vasershtein and A.M. Leonfovich, (1970) "On invarient measures of certain Markov operators, describing a random medium," Problemy Peredachi Informatsii, vol. 6, No. 1, pp. 71-80 (in Russian).

15. I. Csiszár (1975), "I-divergence geometry of probability distributions and minimization problems," Ann.Math. Prob. vol. 3, No. 1, pp. 146-159.

16. R.M. Gray, D. Neuhoff and P.C. Shields (1975) "A generalization of Orstein's distance with applications to Information Theory," Ann.Prob., vol. 3, No. 2, pp. 315-328.

17. V.M. Zolotarev and V.V. Senatov, (1975), "Two-sided Estimates of Lévy's metric," Theory Prob. and Aopl. vol. XX, No. 2, pp. 234-245.

18. D.L. Neuhoff, R.M. Gray and L.D. Davisson (1975) "Fixed Rate Universal Block Source Coding with a Fidelity Criterion," IEEE Trans. Inform. Th. vol. IT-21, No. 5, pp. 511-523.

19. A.D. Wyner, (1975), "The Common Information of Two Dependent Random Variables," IEEE Trans. Inform. Th. vol. IT-21, No. 2, pp. 163-179.

20. S. Cambanis, G. Sinous and W. Strout (1976), "Inequalities for Ek (X,Y) when the Marginals are Fixed," Institute of Statistics mimeo Series no. 1053, University of North Carolina, Chapel Hill.

21. P. Papantoni-Kazakos (1977), "Robustness in Parameter Estimation," IEEE Trans. Inform. Th., to appear.

22. D. Kazakos (1977), "Maximum Linear Discrimination," IEEE Trans. on Systems, Man and Cybermetrics, to appear.

23. R.M. Gray (1976) Private communication.

24. D. Kazakos (1976) Private communication.

25. A. Tchen (1976) Private communication.

Appendix A

For the proof of inequality (3) in section 2, the following theorem by Strassen ([7], Th. 11) is needed.

Theorem I

Given two measures μ_1, μ_2 defined on the separable complete space $(\Omega, \mathcal{Q})$, the inequality

$$\inf \{\epsilon : \mu_1(A) \leq \mu_2(\cup B; B\epsilon\mathcal{Q}, \rho(A,B) \leq \epsilon) + \epsilon \ ; \ \forall A\epsilon\mathcal{Q}\} \leq \zeta \qquad (a_1)$$

is true if and only if there is some joint measure $r(A,B)$ with $\mu_1(A), \mu_2(B)$ marginals such that

$$r(A,B: \rho(A,B) > \zeta) < \zeta \qquad (a_2)$$

Using Theorem I we will prove the following lemna.

Lemna I

If $d_{P,\rho}(\mu_1,\mu_2)$, $d_{V,\rho}(\mu_1,\mu_2)$ are the Prokhorov and Vasershtein distances correspondingly as defined by (1) and (2), then for every $\rho(\cdot,\cdot)$ and measures μ_1, μ_2 defined on $(\Omega, \mathcal{Q})$ and such that $\mu_1(\Omega) = \mu_2(\Omega)$ the following inequality is true:

$$d_{V,\rho}(\mu_1,\mu_2) \geq [d_{P,\rho}(\mu_1,\mu_2)]^2 \qquad (a_3)$$

Proof:

$$d_{P,\rho}(\mu_1,\mu_2) = \inf \{\epsilon : \mu_1(A) \leq \mu_2(\cup B; B\epsilon\mathcal{Q}, \rho(A,B) \leq \epsilon) + \epsilon; \ \forall A\epsilon\mathcal{Q}\} = \epsilon_P$$

For $\delta > o$, there is some $\epsilon_\delta : \epsilon_\rho < \epsilon_\delta \leq \epsilon_\rho + \delta$ and there is some A_o giving:
$\mu_1(A_o) > \mu_2(\cup B; B : B\epsilon\mathcal{Q}, \rho(A,B) \leq \epsilon_\delta) + \epsilon_\delta$.

By theorem 1, there is then no joint measure $r(\cdot,\cdot)$ such that

$$r(A,B:\ \rho(A,B) > \epsilon_\delta) < \epsilon_\delta\ .$$

That is, for every $r(\cdot,\cdot)$ choice it is true that

$$r(A,B:\ \rho(A,B) > \epsilon_\delta) \leq \epsilon_\delta\ .$$

Since $E\{\rho(A,B)\} = \underset{A,B:\rho(A,B) > a}{E_{r(\cdot,\cdot)}}\{\rho(A,B)\} + \underset{A,B:\rho(A,B)\leq a}{E_{r(\cdot,\cdot)}}\{\rho(A,B)\} \geq$

$$\geq a\ r(A,B:\ \rho(A,B) > a)\ ;\ \forall\ a$$

we can obtain that for every $r(\cdot,\cdot)$ choice it is true that

$$E_{r(\cdot,\cdot)}\{\rho(A,B)\} \geq \epsilon_\delta\ r(A,B:\rho(A,B) > \epsilon_\delta) \geq \epsilon_\delta^2\ .$$

Therefore, if by ϵ_v we denote the Vasershtein distance, we have:

$$\epsilon_v = \inf E_{r(\cdot,\cdot)}\{\rho(A,B)\} \geq \epsilon_\delta^2 > \epsilon_p^2\ .$$

every $r(\cdot,\cdot)$ with μ_1,μ_2 marginals.

<u>Proof of Lemna 2</u>

If $\rho(A,B) = \begin{cases} S\ , & A \not\equiv B \\ o\ , & A \equiv B \end{cases}$, then the inequality:

$$\mu_1(A) \leq \mu_2(\cup B\ ;\ B\epsilon\mathcal{Q}\ ,\rho(A,B) \leq \epsilon\ ) + \epsilon$$

is equivalent to

$$\begin{cases} \mu_1(A) \leq \mu_2(A) + \epsilon\ \ ;\ \ \epsilon < S \\ \mu_1(A) \leq \mu_2(\cup B\ ;\ B\epsilon\mathcal{Q}) + \epsilon = \mu_2(\Omega) + \epsilon\ \ ;\ \ \epsilon \geq S \end{cases} \qquad (a_4)$$

If $\mu_1(\Omega) = \mu_2(\Omega) \geq \max\ (\mu_1(A),\ \mu_2(A))\ \ \forall\ A\epsilon\mathcal{Q}$

then the second part of (a_4) is always true (for all $A\epsilon\mathcal{Q}$) because

$$\mu_1(A) \leq \mu_2(\cup B\ ;\ B\epsilon\mathcal{Q}) = \mu_2(\Omega)\ ;\ \forall A\epsilon\mathcal{Q}\ .$$

Thus, the

$$\inf \{\epsilon : \mu_1(A) \leq \mu_2(\cup B; B\epsilon\mathcal{Q}, \rho(A,B) \leq \epsilon) + \epsilon\} \; ; \; \forall A\epsilon\mathcal{Q}\} \qquad (a_5)$$

is equal to S if for every $\epsilon < S$ there is some $A\epsilon\mathcal{Q}$ such that either

$\mu_1(A) > \mu_2(A) + \epsilon$ or $\mu_2(A) > \mu_1(A) + \epsilon$.

That is because then:

$$\sup_{A\epsilon\mathcal{Q}} | \mu_1(A) - \mu_2(A) | > \epsilon \quad , \quad \forall\, \epsilon < S \; , \quad \text{which leads to:}$$

$$\sup_{A\epsilon\mathcal{Q}} | \mu_1(A) - \mu_2(A) | \geq S \; . \quad \text{But since } o \leq \mu_i(A) \leq S \; ; \; \forall A\epsilon\mathcal{Q} \; ;$$

$i = 1,2$, the supremun above can not exceed S and it can only be equal to it.

On the other hand, if for some $\epsilon < S$, the inequalities $\mu_1(A) \leq \mu_2(A) + \epsilon$ are true $\forall\, A\epsilon\mathcal{Q}$, then the infimum in (a_5) becomes equal to:

$$\inf \{\epsilon : \epsilon < S \; , \; \mu_1(A) \leq \mu_2(A) + \epsilon \; , \; \mu_2(A) \leq \mu_1(A) + \epsilon \; ; \; \forall\, A\epsilon\mathcal{Q} \,\}$$

$$= \inf \{\epsilon \; ; \; \epsilon < S \; , \; | \mu_1(A) - \mu_2(A) | \leq \epsilon \; ; \; \forall\, A\epsilon\mathcal{Q}\}$$

$$= \sup_{A\epsilon\mathcal{Q}} | \mu_1(A) - \mu_2(A) | = d_K(\mu_1, \mu_2)$$

The proof is now complete.

<u>Proof of Lemna 3</u>

We have in this case:

$$d_{V,\rho}(\mu_1,\mu_2) = \inf_{r(\cdot,\cdot) \text{ with } \mu_1,\mu_2 \text{ marginals}} r(A,B; A,B \in \mathcal{Q}, A \neq B)$$

$$= \inf_{r(\cdot,\cdot) \text{w.m. } \mu_1,\mu_2} r(A,B; A,B \in \mathcal{Q} \; , \; \rho(A,B) = 1) =$$

$$= \inf_{r(\cdot,\cdot)} r(A,B; A,B \in \mathcal{Q}, \rho(A,B) \geq 1) \qquad (a_6)$$

Let $d_{V,\rho}(\mu_1,\mu_2) = \epsilon_v \leq 1$.

Then, we obtain from (a_6):

$$\inf_{r(\cdot,\cdot)} r(A,B\ ;\ A,B\epsilon\mathcal{Q}\ ,\ \rho(A,B) \geq 1) = \epsilon_v$$

So, for every $\delta > o$ there is some $r(\cdot,\cdot)$ with μ_1,μ_2 marginals such that

$$r(A,B;\ A,B\epsilon\mathcal{Q}\ ,\ \rho(A,B) \geq 1) < \epsilon_v + \delta$$

From theorem I we obtain then:

$$\mu_1(A) \leq \mu_2(\cup B;B\epsilon\mathcal{Q}\ ,\ \rho(A,B) \leq \epsilon_v + \delta) + \epsilon_v + \delta\ ;\ \forall A\epsilon\mathcal{Q} \qquad (a_7)\ .$$

Expression (ϵ_7) is true only for $\delta > o$.

Therefore,

$$\inf\{\epsilon:\mu_1(A) \leq \mu_2(\cup B;B\epsilon\mathcal{Q}\ ,\ \rho(A,B) \leq \epsilon) + \epsilon\ ;\ \forall A\epsilon\mathcal{Q}\} = \epsilon_v \qquad (a_8)$$

But, as shown in the proof of lemna 2,

$$\inf\{\epsilon:\mu_1(A) \leq \mu_2(\cup B;B\epsilon\mathcal{Q}\ ,\ \rho(A,B) \leq \epsilon) + \epsilon\ ;\ \forall A\epsilon\mathcal{Q}\}$$

$$= \sup_{A\epsilon\mathcal{Q}} |\ \mu_1(A) - \mu_2(A)\ | = d_K(\mu_1,\mu_2) \qquad (a_9)$$

From (a_8) and (a_9) we conclude that for the measure $\rho(A,B)$ as expressed by lemna 3, it is true that:

$$d_{v,\rho}(\mu_1,\mu_2) = d_K(\mu_1,\mu_2)$$

<u>Appendix B</u>

<u>Proof of result in 4aI</u>

Denote $x = \dfrac{C_n R_{n2} C'_n}{C_n R_{n1} C'_n}$. Then,

$$\frac{1}{2}\left(\frac{C_n\ R_{n2}\ C_n}{C_n\ R_{n1}\ C'_n}\right)^{\frac{1}{2}} + \frac{1}{2}\left(\frac{C_n\ R_{n1}\ C_n}{C_n\ R_{n2}\ C'_n}\right)^{\frac{1}{2}} = \frac{1}{2}\, x^{\frac{1}{2}} + \frac{1}{2}\, x^{-\frac{1}{2}} = g(x) \qquad (b_1)$$

monotonically decreasing from $x < 1$ to $x = 1$ and monotonically increasing from $x = 1$ to $x > 1$. Therefore, if there are any restrictions on the value

of x , $g(x)$ will assume its maximum for either the minimum or the maximum x allowed. Now, if we define $D = \{d_i \; ; \; i = 1,\cdots,n\} = C_n W$, where $R_{n1} = WW'$, we obtain:

$$x^{-1} = \frac{C_n R_{n2} C_n'}{C_n R_{n1} C_n'} = \sum_{i=1}^{n} \lambda_i \frac{d_i^2}{\sum_{j=1}^{n} d_j^2} \qquad (b_2)$$

where $L = \{\lambda_i\}$; $R_{n2} = WLW'$.

For every i, $o \leq \frac{d_i^2}{\sum_{j=1}^{n} d_j^2} \leq 1$, $\lambda_i > o$ and $\sum_{i=1}^{n} \frac{d_i^2}{\sum_{j=1}^{n} d_j^2} = 1$

Therefore, the maximum value x^{-1} can take is equal to λ_{max} , which is the maximum eigenvalue in L , and this is realized by $D = [o,\cdots o,k,o\cdots o] = C_n W$ (b_3)

Position corresponding to λ_{max}

where k any constant.

From (b_3) we obtain

$$C_n^{\lambda max} = k\,[o,\cdots o,\underset{\uparrow}{1},o\cdot o]\, W^{-1} \qquad (b_4)$$

position of λ_{max}

The minimum value x^{-1} can take is λmin and is similarly leading to:

$$C_n^{\lambda min} = k\,[o,\cdots o,\underset{\uparrow}{1},o,\cdot,o]\, W^{-1} \qquad (b_5)$$

position of λ_{min}

Applying the restriction $C_n R_{n1} C_n' \leq A$ to (b_4) and (b_5) we obtain: $|K| \leq \sqrt{A}$.

<u>Proof for 4aII</u>

Denote $x = \dfrac{C_n R_{n2} C_n'}{C_n R_{n1} C_n'}$

Then

$$d_{Vr}(f_{1t},f_{2t}) = 4 \left| \Phi\left(x\sqrt{\frac{2\ell nx}{x^2-1}}\right) - \Phi\left(\sqrt{\frac{2\ell nx}{x^2-1}}\right) \right|$$

For $x \geq 1$

$\frac{2\ln x}{x^2-1}$ is monotonically decreasing from ∞ to zero with x increasing from $x = 1$ to infinity.

Also, $(x-1)\sqrt{\frac{2\ln x}{x^2-1}}$ increases monotonically for x increasing from $x = 1$ to infinity.

Therefore, $\Phi\left(x\sqrt{\frac{2\ln x}{x^2-1}}\right) - \Phi\left(\sqrt{\frac{2\ln x}{x^2-1}}\right)$ is increasing monotonically with x increasing from $x = 1$ to infinity. Similarly, $\Phi\left(\sqrt{\frac{2\ln x}{x^2-1}}\right) - \Phi\left(x\sqrt{\frac{2\ln x}{x^2-1}}\right)$ is monotonically increasing with x decreasing from $x = 1$ to $-\infty$.

Proof for 4aV:

Denote $x = \frac{C_n R_{n2} C_n'}{C_n R_{n1} C_n'}$; $\sigma_1 = C_n R_{n1} C_n'$

and let us consider that $x < 1$. For $x > 1$ symetric procedure holds.

Then, the Lévy distance in (40) becomes:

$$d_{L,\rho}(f_{1T}, f_{2T}) = \inf\{\epsilon : \Phi(\tfrac{y}{x}) - \Phi\left(y + \frac{\rho^{-1}(\epsilon)}{\sigma_1}\right) \leq \epsilon \; ; \forall y\} \qquad (b_6)$$

if the inequality in (b_6) should be true for every y and for given ϵ, x, it is sufficient that it is true for this y that obtains the $g(y) = \Phi(\frac{y}{x}) - \Phi(y + \frac{\rho^{-1}(\epsilon)}{\sigma_1})$ maximum. Taking the first derivative of $g(y)$ we find

$$g'(y) = \frac{1}{x}\varphi(\tfrac{y}{x}) - \varphi\left(y + \frac{\rho^{-1}(\epsilon)}{\sigma_1}\right) \; ; \text{ where } \varphi(u) = \frac{\exp\{-\frac{u^2}{2}\}}{\sqrt{2\pi}}$$

The derivative $g'(y)$ is nomegative for these y's that satisfy:

$$\ln\frac{1}{x} - \frac{1}{2}\frac{y^2}{x^2} \geq -\frac{1}{2}\left(y + \frac{\rho^{-1}(\epsilon)}{\sigma_1}\right)^2 \qquad (b_7)$$

Or for:

$$\frac{\frac{\rho^{-1}(\epsilon)}{\sigma_1} - \sqrt{\frac{1}{x^2}\left(\frac{\rho^{-1}(\epsilon)}{\sigma_1}\right)^2 + 2\left(\frac{1}{x^2}-1\right)\ln\frac{1}{x}}}{\frac{1}{x^2}-1} \le y \le$$

$$\frac{\frac{\rho^{-1}(\epsilon)}{\sigma_1} + \sqrt{\frac{1}{x^2}\left(\frac{\rho^{-1}(\epsilon)}{\sigma_1}\right)^2 + 2\left(\frac{1}{x^2}-1\right)\ln\frac{1}{x}}}{\frac{1}{x^2}-1} \qquad (b_8)$$

Due to this and the fact that $g(-\infty) = o$, we see that $g(g)$ obtains maximum at

$$y = \frac{\frac{\rho^{-1}(\epsilon)}{\sigma_1} + \sqrt{\frac{1}{x^2}\left(\frac{\rho^{-1}(\epsilon)}{\sigma_1}\right)^2 + 2\left(\frac{1}{x^2}-1\right)\ln\frac{1}{x}}}{\frac{1}{x^2}-1}$$

and for $x < 1$ the distance in (b_6) becomes:

$$d_{L,\rho}(f_{1T},f_{2T}) = \inf\Bigg\{\epsilon:\ \Phi\left(\frac{1}{x}\cdot\frac{\frac{\rho^{-1}(\epsilon)}{\sigma_1} + \sqrt{\frac{1}{x^2}\left(\frac{\rho^{-1}(\epsilon)}{\sigma_1}\right)^2 + 2\left(\frac{1}{x^2}-1\right)\ln\frac{1}{x}}}{\frac{1}{x^2}-1}\right) -$$

$$-\Phi\left(\frac{\frac{1}{x^2}\,\frac{\rho^{-1}(\epsilon)}{\sigma_1} + \sqrt{\frac{1}{x^2}\left(\frac{\rho^{-1}(\epsilon)}{\sigma_1}\right)^2 + 2\left(\frac{1}{x^2}-1\right)\ln\frac{1}{x}}}{\frac{1}{x^2}-1}\right) \le \epsilon \qquad (b_9)$$

If this $x < 1$ that makes the distance in (b_9) maximum can be found, the transformation C_n that obtains this x can be found also as in the cases of Bhattacharyya, I-divergence, variational and Kolmogorov distances. However, the task proves to be such that only numberical methods can approach it.

Proof for 4bV

For the equal covariance case in 4b, the Lévy distance is given by:

$$d_{L,\rho}(f_{1T},f_{2T}) = \inf\{\epsilon: \Phi(x) - \Phi(x + \frac{\rho^{-1}(\epsilon)}{C_n R_n C_n'} + \frac{C_n M_n}{C_n R_n C_n'}) \leq \epsilon ,$$

$$\Phi(x + \frac{C_n M_n}{C_n R_n C_n'}) - \Phi(x + \frac{\rho^{-1}(\epsilon)}{C_n R_n C_n'}) \leq \epsilon \ ; \forall x \} \qquad (b_{10})$$

Denote:

$$g_1(x) = \Phi(x) - \Phi(x + \frac{\rho^{-1}(\epsilon)}{C_n R_n C_n} + \frac{C_n M_n}{C_n R_n C_n}) \qquad (b_{11})$$

$$g_2(x) = \Phi(x + \frac{C_n M_n}{C_n R_n C_n'}) - \Phi(x + \frac{\rho^{-1}(\epsilon)}{C_n R_n C_n'}) \qquad (b_{12})$$

$$\frac{\partial g_1(x)}{x} = g_1'(x) = \varphi(x) - \varphi(x + \frac{\rho^{-1}(\epsilon)}{C_n R_n C_n'} + \frac{C_n M_n}{C_n R_n C_n'}) \qquad (b_{13})$$

$$\frac{\partial g_2(x)}{\partial x} = g_2'(x) = \varphi(x + \frac{C_n M_n}{C_n R_n C_n'}) - \varphi(x + \frac{\rho^{-1}(\epsilon)}{C_n R_n C_n'}) \qquad (b_{14})$$

$g_1'(x)$ is positive for:

$$(2x + \frac{\rho^{-1}(\epsilon)}{C_n R_n C_n'} + \frac{C_n M_n}{C_n R_n C_n'})(\frac{\rho^{-1}(\epsilon)}{C_n R_n C_n'} + \frac{C_n M_n}{C_n R_n C_n'}) > o \qquad (b_{15})$$

$g_2'(x)$ is positive for:

$$(2x + \frac{\rho^{-1}(\epsilon)}{C_n R_n C_n'} + \frac{C_n M_n}{C_n R_n C_n'})(\frac{\rho^{-1}(\epsilon)}{C_n R_n C_n'} - \frac{C_n M_n}{C_n R_n C_n'}) > o \qquad (b_{16})$$

Let ϵ_1 be a candidate for $d_{L,\rho}(f_{iT},f_{2T})$ in (b_{10}).

Then,

i. If $\rho^{-1}(\epsilon_1) \geq | C_n M_n |$

both $g_1'(x)$, $g_2'(x)$ are positive for $x > - \dfrac{\rho^{-1}(\epsilon_1) + C_n M_n}{2 C_n R_n C_n'}$

That means that both $g_1(x)$, $g_2(x)$ obtain maximum at either $x = +\infty$ or $x = -\infty$ and this value is zero. So this case is trivial.

ii. If $\rho^{-1}(\epsilon_1) < C_n M_n$; $\rho^{-1}(\epsilon_1) > 0$

$g_1'(x)$ is positive for $x > - \dfrac{\rho^{-1}(\epsilon_1) + C_n M_n}{2 C_n R_n C_n'}$

and $g_2'(x)$ is positive for : $x < - \dfrac{\rho^{-1}(\epsilon_1) + C_n M_n}{2 C_n R_n C_n'}$

That means that $g_1(x)$ obtains maximum for $x = \pm \infty$ and this maximum is zero (trivial), while $g_2(x)$ obtains maximum for $x = -\dfrac{\rho^{-1}(\epsilon_1) + C_n M_n}{2 C_n R_n C_n'}$

and this maximum is equal to:

$$\Phi\left(- \frac{\rho^{-1}(\epsilon_1)}{2 C_n R_n C_n'}\right) - \Phi\left(- \frac{C_n M_n}{2 C_n R_n C_n'}\right) \qquad (b_{17})$$

and (b_{17}) reduces the search for $d_{L,\rho}(f_{1T}, f_{2T})$ to finding the infinuim ϵ_1 such that the expression in (b_{17}) does not exceed ϵ_1 .

iii. The case $C_n M_n < 0, \rho^{-1}(\epsilon_1) < - C_n M_n$ is synmetric to ii.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1231

4. TITLE: [illegible] DISTANCE MEASURES AND THEIR USE IN FEATURE SELECTION	5. TYPE OF REPORT & PERIOD COVERED: Interim
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): P. Papantoni-Kazakos	8. CONTRACT OR GRANT NUMBER(s): AFOSR 77-3156 new
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Rice University, Department of Electrical Engineering, Houston, TX 77001	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332	12. REPORT DATE: 1976
	13. NUMBER OF PAGES: 36
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Five different distance measures are compared, evaluated and used for data reduction in two separte Gaussian families.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)